

The Military Language Tutor (MILT)

Jonathan D. Kaplan, Mark A. Sabol, and Robert A. Wisher
U.S. Army Research Institute

Advanced Training Methods Research Unit
Robert J. Seidel, Chief

November 1998

Reproduced From
Best Available Copy



U.S. Army Research Institute
for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

19981204 053

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

A Directorate of the U.S. Total Army Personnel Command

**EDGAR M. JOHNSON
Director**

Technical review by

Angelo Mirabella
Bruce Knerr

NOTICES

DISTRIBUTION: This Research Note has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This Research Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this Research Note are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) November 1998		2. REPORT TYPE Final		3. DATES COVERED (from... to) 04-09-92 to 05-05-98	
4. TITLE AND SUBTITLE The Military Language Tutor (MILT)				5a. CONTRACT OR GRANT NUMBER	
				5b. PROGRAM ELEMENT NUMBER 0602785A	
6. AUTHOR(S) Jonathan D. Kaplan, Mark A. Sabol, Robert A. Wisher				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER 2233	
				5e. WORK UNIT NUMBER CO4	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-II 5001 Eisenhower Avenue Alexandria, VA 22333-5600				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Research Note 99-04	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT (Maximum 200 words): MILT is a military foreign language tutor and an authoring system. MILT joins the strengths of previous computer-based approaches to language training with emerging technologies from the fields of computational linguistics, computer science, and electrical engineering to form an innovative, interactive tutor in a Pentium-based laptop computer. The first version of MILT with keyboard input was designed for Spanish and Arabic and can recognize tens of thousands of common words and hundreds of military terms in each of these languages. Its major software engine is a natural language processor (NLP). The goal of the MILT design team was an authoring system which would require no formal external training and which could be learned within four hours by anyone familiar with the Windows environment, even someone with no programming experience, using only documentation and internal MILT help functions. In MILT-DSR(discrete speech recognition), students are given an exercise which allows them to use language production to manipulate a graphics microworld. At Fort Campbell using 5th Special Forces Group personnel a field evaluation was conducted in early June, 1997. For each evaluation, two types of data were collected: (a) student attitudes toward the tutor and (b) instructional effects of the tutor.					
15. SUBJECT TERMS authoring, foreign language, arabic, NLP, speech recognition, intelligent tutor, tutor					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT Unlimited	20. NUMBER OF PAGES 29	21. RESPONSIBLE PERSON (Name and Telephone Number) Jonathan D. Kaplan (703) 617-8828
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified			

Research Note 99-04

The Military Language Tutor (MILT)

Jonathan D. Kaplan, Mark A. Sabol, and Robert A. Wisher
U.S. Army Research Institute

Advanced Training Methods Research Unit
Robert J. Seidel, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

November 1998

Army Project Number
2O262785A790

Personnel Systems and
Performance Technology

Approved for public release; distribution is unlimited.

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducted an advanced language tutor and tutor authoring system development program that was sponsored by the U.S. Army Intelligence Center and School (USAICS), the Department of Defense Office of Special Technology (OST), and the Special Forces Language Office (SOFLO). This resulted in the completion of the MILT-TXT (Military Language Tutor-Text) and MILT-DSR (discrete speech recognition) systems.

At the request of OST and SOFLO, ARI conducted an evaluation of the Arabic version of MILT-DSR with the 5th Group of the Special Forces (SF) at Fort Campbell during the first two weeks in June 1997. The results of this evaluation were briefed to headquarters personnel at TRADOC, OST, USAICS, and SOCOM. Lessons learned in the evaluation were incorporated into MILT-CSR (continuous speech recognition).

ZITA M. SIMUTIS
Technical Director

ACKNOWLEDGEMENTS

The authors acknowledge the cooperation of the DoD Office of Special Technology and the 5th Group of the Special Forces at Fort Campbell, KY. Without their help this project and its evaluation would not have been possible.

THE MILITARY LANGUAGE TUTOR (MILT)

EXECUTIVE SUMMARY

Research Requirement

The acquisition and sustainment of foreign language skills has been an enduring problem for global military readiness. Foreign language skills are notoriously perishable and maintaining proficiency in critical languages is difficult and costly. If a viable approach can be developed, a technological solution for sustainment training at the job-site could be a cost-effective alternative, and offer realistic, interactive practice. The Military Language Tutor (MILT) was designed to fill this need.

Procedure

U.S. The Army Research Institute (ARI) analyzed the state of current technologies and concluded that a technological solution to language sustainment might be possible. As a result, ARI created the MILT contract program. The Military Language Tutor (MILT) is a software system developed to perform two primary functions, authoring and tutoring, required for the development and delivery of computer-based language training (CBLT) programs. MILT, as a tutoring system, has features that distinguish it from other CBT programs developed for specific languages:

MILT is a quasi-intelligent tutoring system (ITS) because its natural language processing (NLP) capability is able to identify specific types of student errors, and because it can alter question sequencing based on student error types and numbers using author-developed models. MILT provides a semi-authorable microworld for practicing language in an intrinsically rewarding, game-like environment. The microworld allows a student to solve an authorable problem by using a target language to manipulate an animated agent searching a semi-authorable series of rooms. Other tutors have used a game-like approach, but MILT is the first tutor to offer a very easy authoring capability so that instructors can create their own microworld exercises.

MILT, as an authoring system, has the capability that allows the instructor or the course administrator to modify the existing tutoring programs in the system, or to develop new ones without requiring any programming skills. This authoring capability can be used for the development of a number of types of instructional programs besides language. However its NLP components are tailored for the development of language tutoring programs.

The first version of MILT with keyboard input was designed for Spanish and Arabic and can recognize tens of thousands of common words and hundreds of military terms in each of these languages. Its major software engine is a natural language processor (NLP). The goal of the MILT design team was an authoring system which would require no formal external training and which could be learned within four hours by anyone familiar with the Windows environment, even someone with no programming experience, using only documentation and internal MILT help functions. In MILT-DSR(discrete speech recognition), students are given an exercise which allows them to use language production to manipulate a graphics microworld. Based on the interest of the Special Operations forces, the Army Research Institute (ARI) and the DoD Office of Special Technology developed a proof of principle version of the Arabic microworld which uses DSR rather than keyboard input to control an animated agent in solving an authored problem. The U.S. Military Academy's Foreign Language Department provided expertise in the development of Arabic acoustic and language models for continuous speech recognition (CSR)

and as part of the project created the first speaker independent Arabic CSR designed for educational purposes.

A pilot test of this proof of principle version of the Arabic MILT was conducted at ARI during April and May of 1997. At Fort Campbell using 5th Special Forces Group personnel a field evaluation was conducted in early June, 1997. For each evaluation, two types of data were collected: (a) student attitudes toward the tutor and (b) instructional effects of the tutor. Information for identifying specific revision needs were collected from students' written responses to explicit questions on surveys and the evaluator's observations of their performance on the tutor. Attitude survey questions concentrated on the 'microworld.' To measure the effects of tutor use on language proficiency, pre and post test measures were taken. Translation tests consisted of English written versions of the 70 microworld command utterances and instructions to speak them in Arabic into a tape recorder. These tape cassettes were later rated by an Arabic linguist who is a native speaker. He rated each utterance of recorded speech (pre and post test) on four dimensions: vocabulary, grammar, pronunciation, and overall fluency without knowing the source of the utterances. The same procedures were used for both pilot and field evaluations.

Findings

On attitudinal measures, the results from both tests were positive. The trainees enjoyed the experience and would like to use the microworld approach in the future. The results from the SF trainees on instructional effectiveness also agreed with the results from the pilot study: all four dimensions of the rated translation score: vocabulary, grammar, pronunciation, and fluency, dramatically increased after an hour's exposure to MILT. The greatest improvement was seen in those participants whose prior skill in Arabic was intermediate; and this group represents the population of users for whom this specific MILT microworld and associated vocabulary were designed. This indicates that a cost-savings, technological aid to the language sustainment problem is now a possibility.

Utilization

MILT is sponsored by the Army Intelligence Center and School, the U.S. Military Academy, and the Special Forces Language Office. It was selected by the Special Forces as the second most important research project being done for them by the Army. The DoD Office of Special Technology has begun a program to combine the key components of the second generation of MILT with those of the CIA sponsored Global Language Authoring System with the intention of creating the most advanced and user-friendly language tutor yet developed.

THE MILITARY LANGUAGE TUTOR (MILT)

CONTENTS	Page
Introduction	1
What is MILT?	1
The Basis of the MILT Approach	3
Design Goals	4
Natural language processing	4
Easy authoring	4
Students in the text microworld	6
MILT and speech recognition	7
Evaluation	9
Field test	10
Summary and Revision Suggestions	14
Speech recognition	15
Graphics	15
Authoring System	16
Next Steps: Dialogue	16
Conclusions	17
References	18

LIST OF FIGURES

Page

FIGURE 1. Features of MILT	2
2. Initial authoring template for creating a microworld	5
3. WYSIWYG dragging of objects into desired location	6
4. Typical MILT-TXT microworld exercise screen with open book about to be read	7
5. MILT-DSR discrete speech microworld	8
6. Relationship between DLPT score and attitude toward using MILT microworld	11
7. Relationship between DLPT score and attitude toward MILT microworld approach	11
8. Relationship between DLPT score and attitude toward advanced MILT	12
9. Pre and posttest means of "overall" ratings of utterances	12
10. Effect of MILT on vocabulary performance as a function of an individual's self-report proficiency	13
11. Effect of MILT on grammar performance as a function of an individual's self-report proficiency	13
12. Effect of MILT on pronunciation performance as a function of an individual's self-report proficiency	14
13. Effect of MILT on fluency as a function of an individual's self-report proficiency	14
14. Example of computer-generated dialogue partner graphic	17

The Military Language Tutor (MILT)

Introduction

The acquisition and sustainment of foreign language skills has been an enduring problem for the Army. To maintain global readiness in the post-cold war world, commanders need personnel who can use critical languages on short notice for a variety of missions. Coordinating actions with allies, understanding intercepted messages from adversaries, and interrogating prisoners, for example, demand high levels of proficiency in listening, reading, and speaking. Indeed, the Army requires soldiers in language-dependent specialties to demonstrate proficiency in both reading and listening on a regular basis.

However, foreign language skills are notoriously perishable. Soldiers' language proficiency, especially speaking skills, have been found to decay significantly almost immediately after their language schooling, as they finish training in their military specialties (Lett & O'Mara, 1990). Yet the traditional means of maintaining language proficiency are difficult and costly, typically requiring intensive instructor time or expensive stays overseas for "immersion" programs. A further complication is that the 11,000 military linguists who would need sustainment training on a recurring basis are scattered about the world, making centralized instruction impractical.

This is clearly an opportunity for computer-assisted training; a technological solution for sustainment training could be a highly cost-effective alternative. The problem is finding an effective approach. Although conventional computer-based language training has been used effectively in promoting vocabulary acquisition (e.g., Berleant, Shi, Wei, Viswanathan, Chai, Majid, Qu, and Sunkara, 1997) and other component language skills amenable to routine, drill-and-practice training procedures, it does not offer -- as does face-to-face dialogue -- the sort of extended, realistic, interaction that has long been considered necessary for true language acquisition (Ellis, 1986; Douglas, 1995; Hamburger, 1995). This deficit has been due in turn to the absence of computer technology capable of sustaining such life-like interactions. The Military Language Tutor (MILT) was designed to correct this deficiency. This paper describes the evolution, field testing, and future prospects for MILT.

What is MILT?

MILT is a unique language technology, designed to be a cost-effective, portable, on-demand alternative to instructor-intensive language training for skill sustainment. The current tutor focuses on sustainment training for Arabic and Spanish. Its key features are portability to the job site and an easy-to-use authoring capability for building instruction in animated graphics environments. It was developed by the U.S. Army Research Institute (ARI), with the support of the Defense Advanced Research Projects Agency and the sponsorship of the DoD Office of Special Technology and the U.S. Army Intelligence Center, Fort Huachuca.

The MILT system draws on lessons learned from earlier ARI language tutor projects (Kreyer & Criswell, 1995; Sams, 1995), although these did not involve speech recognition. Capable of running on a Pentium-based laptop computer, MILT represents a revolutionary approach, combining an easy-to-use authoring system with an effective, interactive computer-tutor. The

authoring system makes it possible for instructors in the field to adapt the tutor to special needs. MILT accomplished this by joining the strengths of previous computer-based approaches to language training with emerging technologies from the fields of computational linguistics, computer science, and electrical engineering. Figure 1 below provides a snapshot of the spectrum of technologies integrated into MILT, proceeding from the more mature technologies on the left to the new, emerging technologies being explored through this research on the right. MILT requires a Pentium-based PC with 32MB of RAM, SVGA graphics, and a Sound Blaster compatible audio card along with a microphone for voice entry.

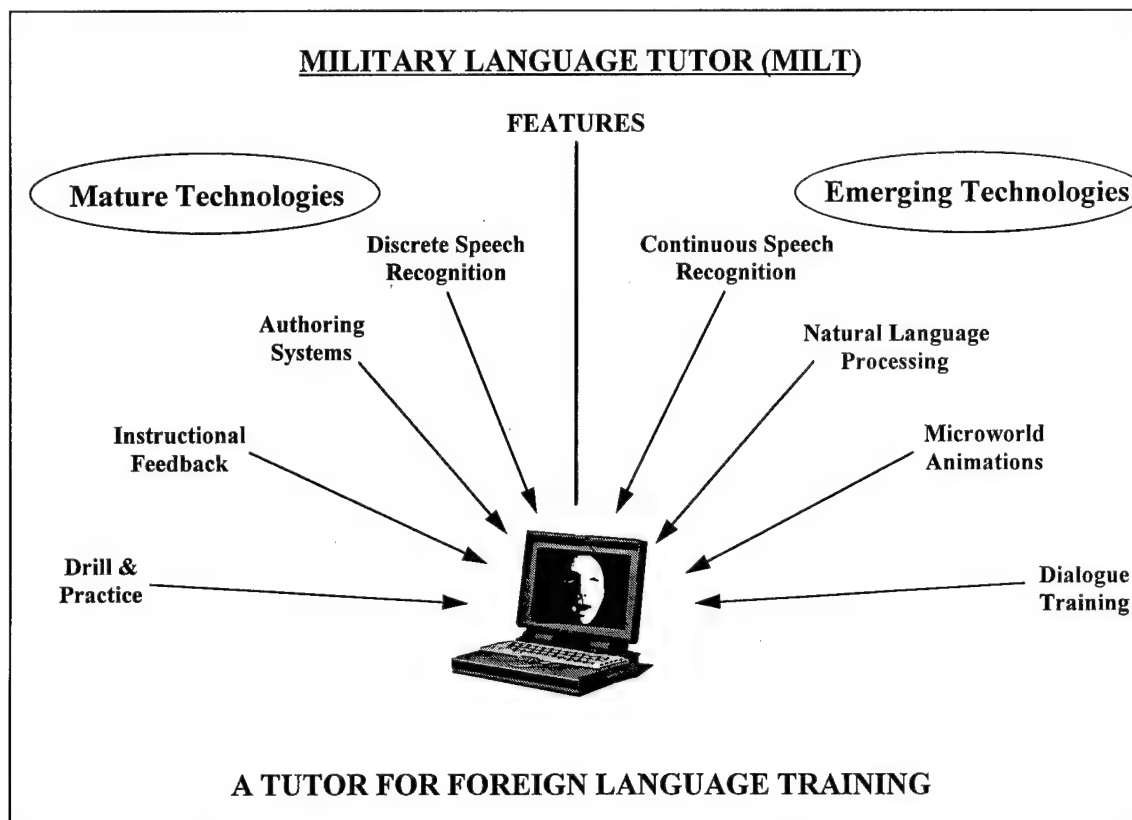


Figure 1. Features of MILT

From its inception, the MILT project was designed to investigate the feasibility of using natural language processing (NLP) software to identify categories of linguistic errors; identified errors could then be used to customize training by way of a dialogue between soldier and computer. MILT is capable of producing all the conventional exercise types and an authorable animated immersion exercise type called a microworld. This exercise immerses the soldier in a mission-relevant world and allows the soldier to control that world by interacting with an animated agent. That interaction requires the soldier to speak or type commands, manipulating the agent as it accomplishes tasks such as searching an office for information in the form of text, sound, and still and moving images. Because such exercises satisfy the primary learning conditions for sustainment: practice, interaction, and feedback, as discussed below, they are expected to be highly motivating. (For a thorough discussion of the general microworld concept, see Schoelles and Hamburger, 1996.) It is the MILT microworld that is the focus of this paper.

The Basis of the MILT Approach

In training literature, "overlearning" is said to occur when training continues beyond simple "mastery," the point when the trainee first performs correctly; the effect of "overlearning" on knowledge and skill retention has been well documented in the research literature (Loftus, 1985). The more soldiers apply language (and other cognitive) skills - beyond the point of initial mastery - the longer they will retain those skills. The importance of "overlearning" grows when stress is added to the situation, as stress has been shown to reduce performance on nearly all skills, particularly those that might be marginal (Driskell, Willis, & Cooper, 1992). Since the purpose of language training in the military is to enable realistic performance in situations that are frequently stressful, and to support retention of these skills over periods of non-use, it is desirable that Army language skills be "overlearned." Therefore, the key to language sustainment is to design a tutor that so interacts with soldiers as to encourage practice beyond the point of mastery.

But how does one attain a state of "overlearning"? It is well established in behavioral research that extrinsic rewards improve learning (Skinner, 1938; Miller, 1963). Extrinsic reward, such as getting a high score on a performance test or the proverbial "rat getting a piece of cheese at the end of the maze," could maintain the behavior involved in foreign language learning beyond mastery. There have, indeed, been attempts to apply such concepts as practice and reinforcement from behavioral psychology to language learning (Skinner, 1957). However, in more recent, cognitive formulations of learning, emphasis is placed on concepts such as feedback, knowledge of results, and intrinsic reinforcement (Levine, 1975; Schmidt & Bjork, 1992). Extrinsic reward does not explain such behavior as people endlessly playing computer games that they do not win, artists painting without hope of selling, or people doing crossword puzzles without expecting to finish them. In these behaviors, people are motivated intrinsically. The behaviors that people find intrinsically motivating appear to be quite diverse, yet they share an underlying commonality. These common elements include 1) exercising control over the elements of the environment, 2) engaging in problem solving, and 3) seeing visible progress (Malone, 1981).

In the case of MILT, we have attempted to make use of intrinsic motivation by providing a game-like microworld in which students use and practice a target language to manipulate an on-screen "microworld" environment as a means of solving a problem. In this way, students can exercise control over a synthetic environment, engage in problem solving, see visible progress, and practice language as a byproduct. We assume that they will find this intrinsically rewarding and will thus be motivated to "overlearn." Furthermore, the types of exercises built into MILT enable realistic communicative interactions. The benefits of such a communicative approach to language teaching have been discussed by Douglas (1995), whose LingWorlds tutor provides a precedent for using language to solve non-language problems. These theoretical considerations have been the impetus for the design and development of MILT. For a more complete discussion of these learning theory issues, see Kaplan and Holland (1995) and Oxford (1995).

The remainder of this report elaborates the key features of MILT and describes the progression of MILT from a text-based to a speech-based training system. Details of the field evaluation at Fort Campbell are provided and an overview of future technological directions for improving MILT is described.

Design Goals

Natural language processing

The first version of MILT (MILT-TXT) is a keyboard entry, text output system that was designed for Spanish and Arabic and can recognize tens of thousands of common words and hundreds of military terms in each of these languages. The large lexicons internal to MILT in these languages, especially Arabic, were constructed from pre-existing lexical resources (Dorr, Garman, & Weinberg, 1994). MILT's major software engine is a natural language processing (NLP) engine that uses parsing mechanisms described by Weinberg, Garman, Martin, and Merlo (1995) and that incorporates semantic and dialogue analysis components discussed by Dorr, Hendler, Blanksteen, and Migdaloff (1995). The NLP concept, adapted from the software used to do machine translation, can analyze language form as well as meaning.

By employing NLP, MILT can address language production skills, in contrast to the typical computer-based language tutor, which focuses on listening and reading skills. Syntactic parsers, semantic analyzers, and artificial intelligence components allow MILT to recognize language and to generate its end of a dialogue. As a result, users of MILT can relatively, freely create language as in the real world, and the system responds by carrying out commands or answering questions. Likewise, tutoring systems can typically tell the difference between right and wrong answers, but they do not know why an answer is wrong. The NLP component enables MILT to identify the type of mistake that was made. This ability allows MILT to adapt a lesson as it is going on. MILT can give additional help to poor students and advance good students to more challenging exercises.

Easy authoring

MILT was designed to be a no-compromise, easy authoring system. That is, all design decisions were made to favor ease in learning and using the system. The goal of the MILT design team was an authoring system which would require no formal external training and which could be learned within four hours by anyone familiar with the Windows environment, even someone with no programming experience, using only documentation and internal MILT help functions. Most of the MILT authoring system is, therefore, based on the template approach. In this approach, the course designer selects a type of question, and the system displays a template for that question type. The template has windows or fields for various kinds of designer input, and buttons or check boxes by which the designer picks from alternatives such as sound versus text presentation of question (see Figure 2). The alternative approach of authoring by writing commands in a computer language (or higher level scripting language) was avoided; though it might provide greater flexibility and a higher level of control, this approach would require much more learning and practice time to master without providing meaningfully greater functionality.

MicroWorld with Voice Recognition

Exercise # 1 Exercise Name: Unnamed

Introductory Directions Type:

☒ Text Question:

☐ Oral Question File Name: Find... Record

Answer:

Number of Tries Allowed: 1

Parser Feedback More...

Edit Speech Text... Edit Scene View Exercise

OK Cancel Help

Figure 2. Initial authoring template for creating a microworld

MILT's microworld authoring is actually a compromise between the template and the WYSIWYG (what-you-see-is-what-you-get) approaches. This section will present a general description of this compromise approach. The next section ("Students in the text microworld") will provide a more specific example of the implementation of this approach. In the typical WYSIWYG approach, authors have buttons, dials, text fields, video fields, etc., which they drag to the locations in which they will appear to the student users. Once they have created the student interface in this "drag and drop" manner, the authors have to tell each element of that interface how to behave. Typically, this is accomplished by writing computer code (or a higher level scripting language) for each element. In MILT's compromise approach conventional exercises are developed using templates, and a template is used to select among alternative room backgrounds and various objects. Then the microworld author drags the actual objects to their desired location on the room background in the traditional WYSIWYG way. To be more specific, the author can start the microworld template and push the appropriate controls to create a room. The author uses the template to select among room backgrounds and objects. The WYSIWYG approach is used to position these objects in the room (see Figure 3). If the author wants to create a second room, he selects and places a door object which serves as a link to a new room. The author never writes computer code to create or control the microworld. Even animated objects in the microworld come with built-in attributes that control their behavior without further work by the author.

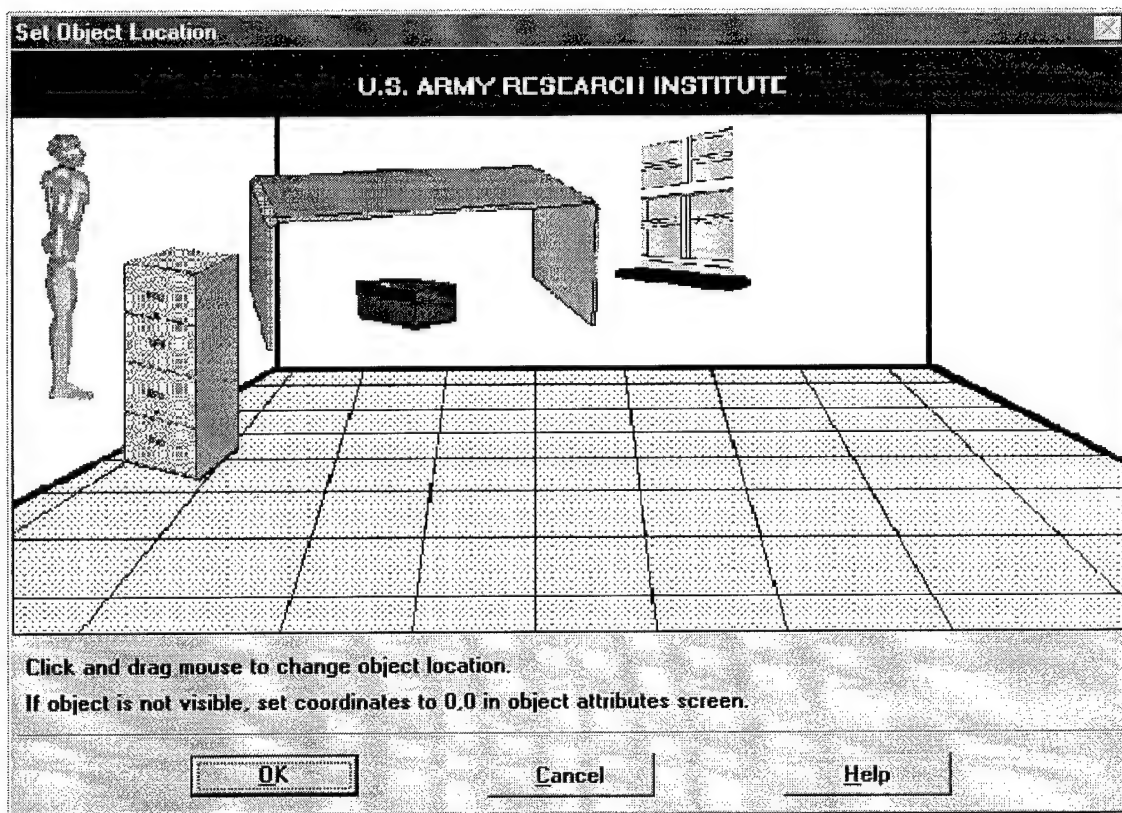


Figure 3. WYSIWYG dragging of objects into desired location

Students in the text microworld

In the MILT microworld, students are given an exercise that allows them to use language production to manipulate a graphics microworld. The manipulation of the graphics by successful language use is a version of controlling the environment and is thought to be intrinsically rewarding. In addition, the microworld is set in a problem-solving domain, and the problem solving itself should be intrinsically rewarding. The question of sources of intrinsic reward is discussed on page three of this paper.

In practice, with the text version of the microworld found in MILT-TXT students are able to see a graphic scene and enter a natural language command in the target language using the keyboard. That entry might contain syntactic, semantic, or factual errors. The NLP engine will analyze the entry, convert it to an interlingual representation, and activate the appropriate animated graphics. An interlingual representation is a symbolic (language independent) representation. If the entry is correct in all respects, the student is rewarded by the execution of his command in the microworld. If it is incorrect in a way that the NLP and animation can deal with, the student sees the results of his language errors taking place in the microworld. If the entry is incorrect in ways that can be identified but not displayed, the tutor tells the student about his errors and invites him to continue.

This microworld is partially authorable and includes the ability to create task-oriented problem solving scenarios with specified goals. The advantages of an animated microworld over typical video training include: greatly shortened production time, greatly reduced production

cost, much higher levels of control for authors, and much higher levels of interactivity for trainees. Figure 4 shows a microworld screen that illustrates text input controlling graphics.

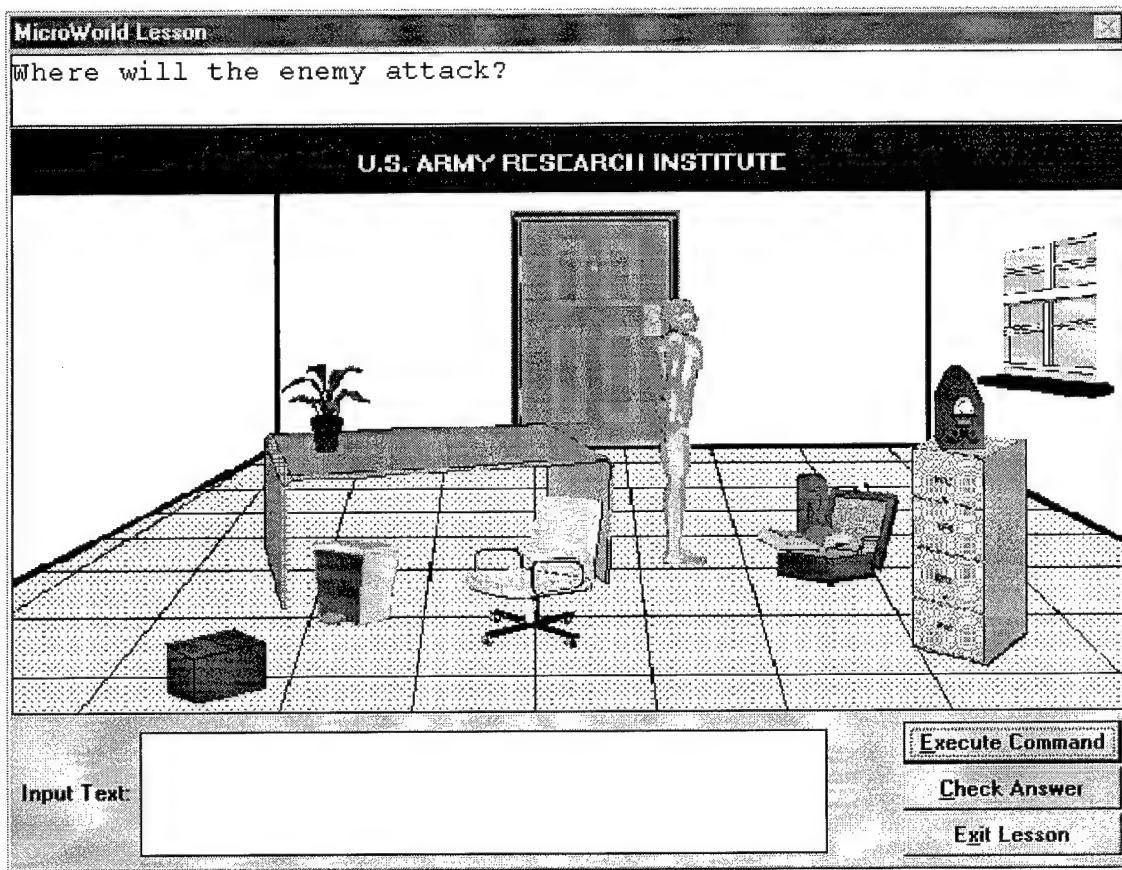


Figure 4. Typical MILT-TXT Microworld exercise screen with open book about to be read.

MILT and speech recognition

There are two general categories of speech recognition systems—discrete speech recognition (DSR) and continuous speech recognition (CSR). Discrete speech recognition was used in MILT-TXT, because continuous speech was not available for use in a Windows or DOS environment at the time of the research. (a screen from the MILT-DSR microworld is shown in Figure 5). This section describes the differences between these two varieties of speech recognition.



Figure 5. MILT-DSR discrete speech microworld.

DSRs recognize individual, separated words, that is, words bounded by pauses. However, people do not normally speak with separations between words. Fortunately, DSRs can recognize quite long individual words; this attribute allows designers to trick DSRs into recognizing short sentences or phrases by handling them as if they were long words. When this trick is employed, however, DSRs view the sentence as a word, and speakers have to use that exact sentence to be recognized. Any alteration of any of the individual words in the sentence may convince DSRs that a different “word” has been uttered and it will not be recognized. This aspect of the “trick” approach leads authors to write potential utterances on the computer screen, allowing users to select and read only one of these. In this way, no utterance will be spoken which the DSR does not already know.

There are two problems with this approach when used in a language tutor—authors have to foresee all likely statements that trainees will want to make and program the written versions of these utterances in advance; and practicing reading is significantly different from practicing spontaneous speech. There are also two positive features of using a DSR in a tutor: 1) it is much less expensive to create the kinds of data models used by DSRs than the kinds required by CSRs, and 2) as long as trainees are reading rather than spontaneously speaking, there is quite a high probability that their utterances will be recognized. It is true that the DSR approach does not simulate real conversation, but it does allow trainees to improve their pronunciation.

The second kind of speech recognizer, CSRs, can recognize sentences consisting of words that run into each other in the way of normal speech. Pauses between words are not required for recognition. Since real individual words are being recognized, it is not necessary to read sentences perfectly, as in the case of DSRs. They can be spontaneously spoken. This eliminates the DSR's problem of having to determine exactly which utterances to write on the screen (to be read) at any given time.

It is clear that it is CSR technology that would permit practice of realistic, spontaneous speech. However there are three significant and currently unsolved problems with this use of CSR: 1) the larger the number of words that can be recognized, the less accurate the recognition; 2) creation or expansion of CSR models requires a speech scientist; and 3) creation of a required acoustic model for a given language typically is very expensive.

Based on the interest of the Special Operations forces (SOF), ARI and the DoD Office of Special Technology (OST) developed a proof of principle version of the Arabic microworld which uses speech rather than keyboard input to control an animated agent in solving an authored problem. The speech is recognized using a Dragon discrete speech recognizer utilizing a relatively small speech model for modern standard Arabic. This version of the MILT-DSR microworld is controlled by 70 spoken Arabic sentences.

Evaluation

A pilot test of this proof of principle version of the Arabic Military Language Tutor (MILT) was conducted at The Army Research Institute (ARI) during April and May of 1997 using eight university students enrolled in Arabic classes as subjects. The purposes of this evaluation were to estimate the overall acceptance level of the proof of principle Arabic speech recognition microworld component in (MILT), and to find specific revision needs before conducting a formal evaluation. The same procedures were used for both evaluation studies.

For estimating the overall acceptance level of the tutor, two types of data were collected: 1) instructional effects of the tutor, and 2) student attitudes toward the tutor. Information for identifying specific revision needs was collected from students' responses to survey questions given during and after their study of the tutoring program, their comments on specific features of the tutor, and the evaluator's observations of their performance on the tutor. Students' responses to the survey questions were also used to assess their attitudes toward the tutor. Although MILT contains all the major classes of exercise types, the attitude survey questions concentrated on the microworld.

To measure the effects of tutor use on language proficiency, pretest and posttest measures were taken. These consisted of giving students a written English version of the 70 microworld command utterances and asking them to speak them in Arabic into a tape recorder. Speech data were collected on two cassettes per student, one for pretest utterances and the other for posttest utterances. The sentences were formed from terms (objects and actions) employed in the MILT microworld. After the completion of data collection, the tapes were rated by a native speaker of Arabic who is also a linguist. For each student, he did blind ratings of the recorded speech (pretest and posttest) on four dimensions: vocabulary, grammar, pronunciation, and fluency. Each utterance was assigned a score ranging from 0 (entirely negative or no attempt) through 3 (acceptable) to 5 (excellent, perfect) on each of these dimensions.

The pilot data showed improved Arabic usage by the college students for more than half of the 70 utterances on each of the four rated dimensions. The attitude survey, focused on reactions to the microworld game approach, indicated highly favorable opinions. Only minor technical revisions of MILT were deemed necessary before the formal field test.

Field Test

The formal study was conducted with 16 members of the 5th group of the Special Forces (SF) at Fort Campbell during the first two weeks in June, 1997. The resulting attitudinal data agree, in general, with the pilot test results: the SF trainees liked the experience and would like to use the microworld approach in the future. The data are presented below as a function of soldiers' self-report of their most recent scores on the Defense Language Proficiency Test (DLPT). This test, required periodically of SF soldiers, is administered by the Defense Language Institute (DLI); possible ratings on each portion of the test (reading, speaking, and listening) are 0, 0+, 1, 1+, 2, 2+, 3, 3+, 4, 4+, and 5.

Because ratings for speaking proficiency were not reported by all subjects, the proficiency of the soldiers was measured based on reported DLI ratings for reading and listening. The average of each soldier's reading and listening scores was calculated. Half the soldiers scored 1 plus or higher on these average ratings and half the soldiers scored below this level. Three subgroups of soldiers were then formed as follows: those whose average DLPT score was less than or equal to 1, those whose average DLPT score was greater than 1 but less than 2, and those whose average DLPT score was 2 or higher. Each figure below shows the distributions of responses to attitude survey questions separately for these three subgroups of subjects. Figure 6 shows how responses within the three groups were distributed in answering the question, "Did you find using the MILT microworld to be enjoyable?" Figure 7 shows the distributions in response to the question, "What is your opinion of the general MILT game approach to language training?" Figure 8 gives the distributions of responses to the question, "Would you regularly use MILT in your own language learning?"

These figures show that reactions to the microworld approach were positive for most soldiers. A majority of trainees, regardless of proficiency, thought MILT was enjoyable to use (Figure 6) and a "very good" or "excellent" idea (Figure 7). However, those who knew less to start with tended to think more highly of MILT than did those who knew more. Those who had low to moderate proficiency beforehand were more positive that they would use MILT for their own language practice (Figure 8). Most subjects reported that the use of MILT was an enjoyable experience and that they would like to use longer and more advanced versions of the microworld as part of their regular language sustainment training.

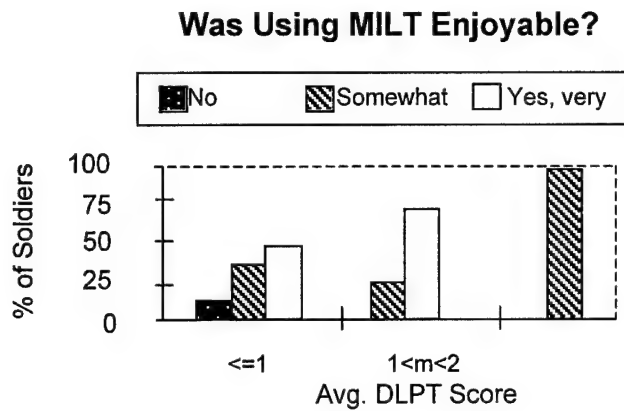


Figure 6. Relationship between DLPT score and attitude toward using MILT microworld.

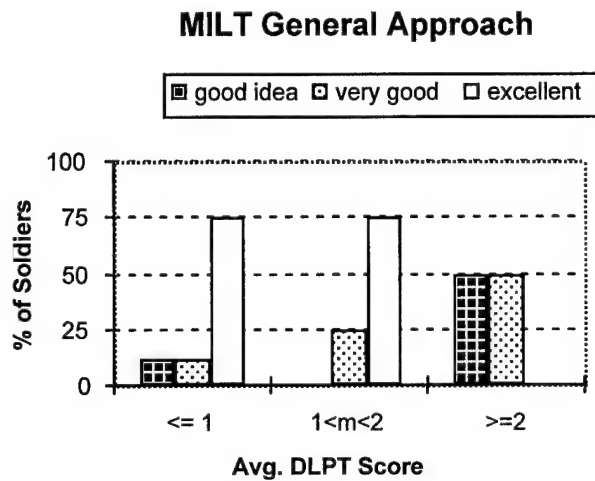


Figure 7. Relationship between DLPT score and attitude toward MILT microworld approach.

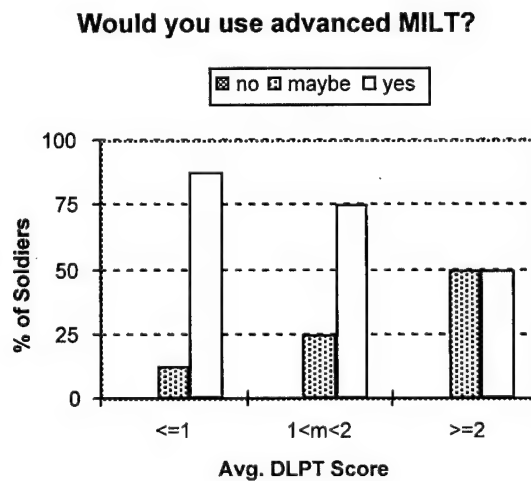


Figure 8. Relationship between DLPT score and attitude toward advanced MILT

The field test trainees did report difficulty getting the tutor to recognizing the utterances they gave; this reaction might indicate a required modification in MILT's speech recognition component. Subjects in the pilot study had also reported this difficulty in recognition by MILT when they first began to use the microworld. Indeed, a modification to alleviate this problem was the principal technical revision of MILT that resulted from the pilot study. However, one interpretation of the fact that the same reported difficulties occurred in the field test calls into question whether the "problem" is a result of a faulty recognizer. Both groups of students felt the tutor had less trouble recognizing their utterances at the later stages of the exercise than at the beginning; this suggests improvement in pronunciation, exactly the effect MILT is intended to produce. Perhaps the most important motivational data collected at Fort Campbell was the fact that the students felt the microworld caused them to practice utterances much more than they normally would have done. This may well have resulted in their improved pronunciation over time. This idea brings us to the data from the ratings of taped utterances.

An overview of the rating data shows clear improvement. Averaged across all 70 sentences and all 16 participants, the mean ratings from the translation test given before exposure to MILT were 1.26, 0.9, 0.9, and 0.8 for vocabulary, grammar, pronunciation, and fluency, respectively. The ratings for translations produced by the same participants after using MILT averaged 2.2, 1.9, 1.5, and 1.4. All four dimensions nearly doubled.

For statistical analysis, the four ratings for each sentence were averaged into an "overall" rating, and each participant was treated as a full experiment. Samples consisted of 70 difference scores, the differences in overall rating for all the sentences produced before and after using MILT. Figure 9 shows, for each subject, the means of the "overall" ratings for the 70 utterances prior to and after MILT training. All but two subjects showed statistically reliable improvement ($p < .05$) in overall translating skill as a result of using MILT for one hour.

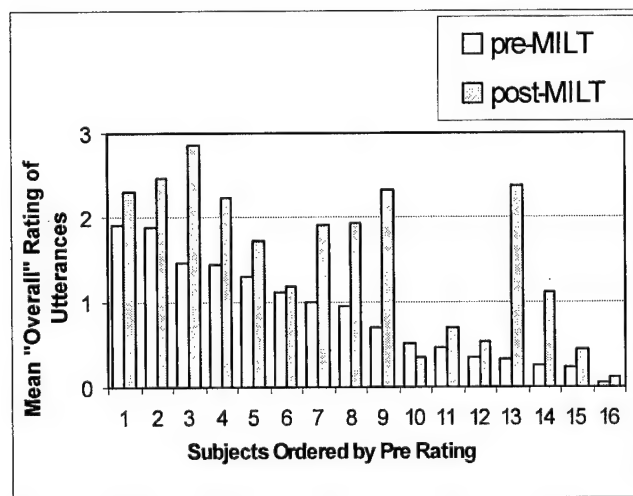


Figure 9. Pre and posttest means of "overall" ratings of utterances.

The rating data were collapsed the other way to show that the improvement occurred for all four rating dimensions. That is, the ratings for each dimension were combined across subjects with similar language proficiency. Figures 10-13 show vocabulary, grammar, pronunciation, and

fluency performance before and after MILT as a function of an individual's self-report proficiency in Arabic. As above, the average proficiency (DPLT) measure is the mean of individual's self-report of their most recently measured proficiency level in reading and listening.

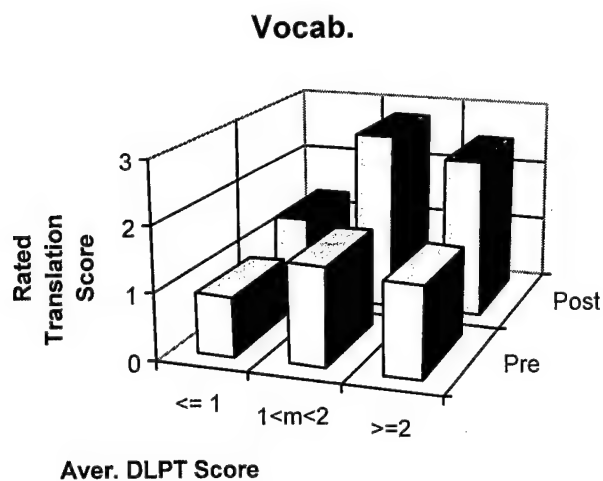


Figure 10. Effect of MILT on vocabulary performance as a function of an individual's self-report proficiency

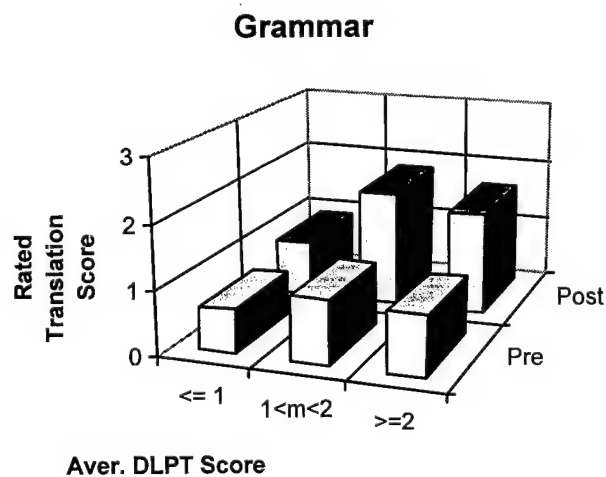


Figure 11. Effect of MILT on grammar performance as a function of an individual's self-report proficiency.

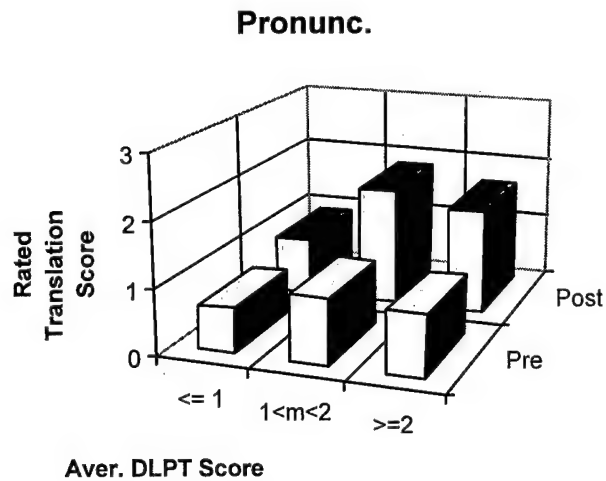


Figure 12. Effect of MILT on pronunciation performance as a function of an individual's self-report proficiency.

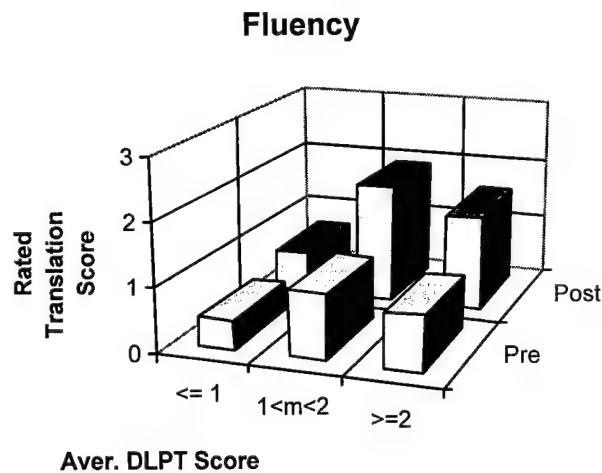


Figure 13. Effect of MILT on fluency as a function of an individual's self-report proficiency.

As in the pilot test, improvement was not uniform. The greatest improvement is seen in those participants whose prior proficiency in Arabic was intermediate. This finding was especially encouraging, since the intermediate group represents the population of users for which this specific MILT microworld and associated vocabulary were designed.

Summary and Revision Suggestions

The overall results of the pilot test and field evaluation showed that one hour's use of the MILT Arabic tutor significantly improved students' Arabic proficiency on four dimensions of language usage--vocabulary, grammar, pronunciation, and overall fluency--and that students' attitudes toward the tutor were favorable. They particularly liked the game approach taken in the

microworld. Suggested revisions derived from the students' responses to survey questions are being carried out in the next phase--the speech recognition version of the MILT program.

Speech recognition

In the MILT-SR program, ARI is investigating both dynamic manipulation of speech recognizer "confidence levels" and the mathematical development of intermediate acoustic models. In the case of DSRs, ARI is developing a process that will start trainees at a "confidence level" that will usually result in MILT recognizing their speech. After a certain number of correct utterances from the student, the DSR will automatically raise the "confidence level," the degree of similarity to the native speaker data that MILT requires before it will recognize an utterance. During the process, if trainees make a significant number of speech errors, the system will automatically alter itself so as to allow greater difference with the native speaker data. This process of raising and lowering the required similarity level will continue throughout the trainee's interaction with MILT-SR, with an ultimate goal of raising it to the native speaker level without significant errors being made. This approach has a liability, however: the problem of false recognitions—if students see the tutor recognizing as correct, utterances they know to be far from perfect, their confidence in the tutor may be weakened.

In the case of CSRs, ARI and the U.S. Military Academy are jointly engaged in a project to develop Arabic intermediate, student acoustic models as well as an Arabic native speaker acoustic model. If this sub-project is successfully developed and incorporated into MILT rather than raising and lowering the acoustic model's confidence level, the confidence level will always be high, but there will be more than one set of acoustic model data. At each level of student proficiency, the system will keep track of the number of errors the student makes and use this information to move the student from one intermediate model to another always keeping the confidence level for a given acoustic model high. Ultimately the student will be moved to a native speaker model. This approach would eliminate much of the false recognition problem.

Using a CSR rather than a DSR should eliminate one pedagogical problem by allowing the elimination of the written versions of allowable utterances in favor of somewhat constrained spontaneous speech. This constraining will be accomplished by allowing students to access all the various objects on the microworld at any time to determine what they are and what can be done with them in the target language. Furthermore, the use of a state-of-the-art CSR should improve recognition accuracy.

Graphics

There is little doubt that the graphics and animation in the current, two dimensional MILT microworld superficially are less impressive than those found in modern game programs, even though below the surface their object orientation and easy authorability make them significantly more sophisticated than current commercial games. However, the difference between what students see in the two dimensional microworld and what they are accustomed to in the commercial world tends to have a negative effect on their attitude. Also, since first person, three dimensional animation appears to be more compelling than the two dimensional, remote agent type used in the current microworld, it is likely that more practice time will be devoted to it. The availability of Windows95 makes such three-dimensional, first-person animation practical for such a project. The microworld is, therefore, being rewritten as a three-dimensional, multi-room environment in the style of virtual presence. The students will experience themselves moving in

a three dimensional world. When they give a command, rather than seeing an animated agent move against a fixed background, they will see the three dimensional background move as if they were moving in that world and seeing the results. The product resulting from this change is expected to be visually impressive.

Authoring System

Changes also will be made in the next version of MILT authoring system. The MILT-CSR microworld will be authorable from a library of objects that can be chosen by authors. The scenario backdrop will be able to be changed as well. A greatly enlarged set of objects will be made available and they will be three dimensional and rotatable. Readable objects like books, newspaper, notebooks, envelopes and letters will be able to be rewritten. Objects that produce sound, such as radios and tape recorders, will be able to produce speech via recordable WAV files. Objects that produce video, such as televisions and VCRs, will be able to play digitized video files. As noted above, MILT-CSR will differ from MILT-TXT by replacing the animated agent with a first-person virtual presence view. Trainees will also be able to pick up and carry objects from one room to another and inventory them on demand. Authors easily will be able to create multiple interconnected rooms. The number of rooms and the size of rooms or outside scenes will be limited only by disk storage.

Next Step: Dialogue

As an outgrowth of the advances made in the MILT program, ARI has begun a project to develop an authoring system for dialogue-based training and expert assistance. Because this project is based on the development of an authorable knowledge base, non-programmers who understand new training or assistance domains and have no specific programming skills will be able to create new dialogue-based tutorials and expert help. The authorable dialogue capability will be a text input-text output system.

It is hoped that in a future phase, the authorable dialogue capability will be embedded in a graphics-based tutor/expert system with full speech recognition and generation capabilities. This will allow soldiers to use the assets of a virtual environment while interacting with a computerized dialogue partner. Because the dialogue partner will be generating his end of the conversation rather than speaking prerecorded material, the graphic portrayal of that partner will have to be generated animation rather than video or typical frame-based animation (see Figure 14).

The current project allows us to address several interesting and difficult research problems. One is the problem of knowledge base creation. A dialogue must be about something, the content of the knowledge domain. If a computer program is to be capable of participating in a dialogue, that potential dialogue content must be organized in a form which is stored somewhere and which can be accessed when needed. Such accessible, organized storage of content is called a knowledge base. Computer knowledge bases have existed for some time, but only in complex form. If new ones need to be created or old ones need to be changed, experts in the knowledge domain must consult computer programmers or acquire complex programming skills themselves. This raises the time and expense of such creation, often to a prohibitive level. Even if the knowledge base is built, the programmer constitutes a filter between the domain expert and the final product, which may, as a result, be faulty. We hope, therefore, to develop a dialogue

authoring system in a direction that will allow nonprogramming domain experts to create knowledge bases with the minimum of training and effort.

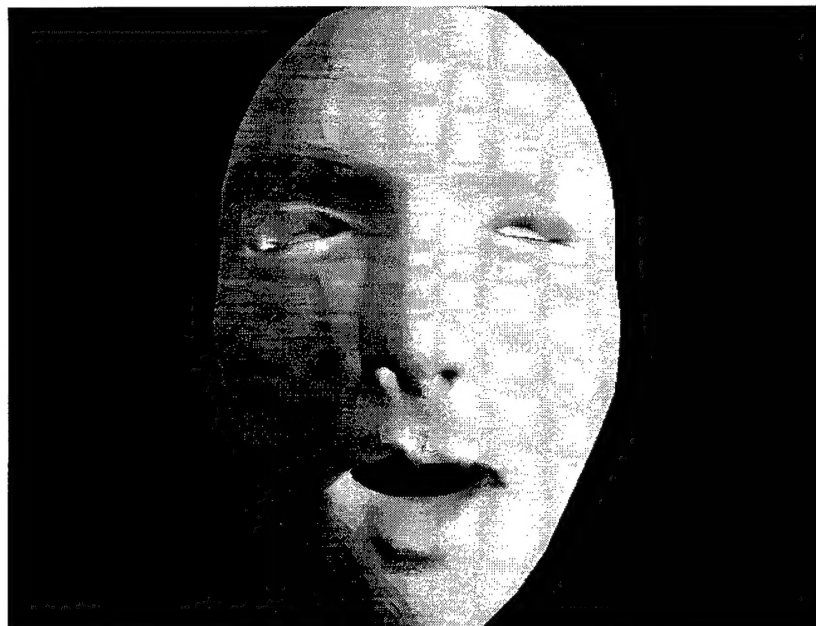


Figure 14. Example of computer-generated dialogue partner graphic (from University of British Columbia).

A second research problem is comprehension: how to get a tutoring or expert help system to understand what its user is trying to communicate so that a dialogue can be initiated. There are three approaches to the comprehension problem: natural language processing (NLP) on its own, continuous speech recognition (CSR) on its own, and a combination of the two. The effectiveness of NLP and CSR alone and in various combinations is not known. Each combination is likely to have its own tradeoff of expense and accuracy of results. Research is needed, therefore, to find the best compromise solution.

Conclusions

From the beginning of the MILT project, practical product development and user needs have guided the research. Earlier versions have already been delivered to military users for implementation. The tutor-dialogue-authoring system that is emerging at the end of this research program will allow instructors to develop lessons and expert help in a wide variety of domains. The resulting system will be a significant advance in state-of-the-art dialogue-based authoring tools and tutors

References

- Dorr, B., Garman, J., & Weinberg, A. (1994). From subcategorization frames to thematic roles: Building Arabic lexical entries for foreign language tutoring.
- Dorr, B., Hendler, J., Blanksteen, S., & Migdaloff, B. (1995). On Beyond syntax: Use of lexical conceptual structure for intelligent tutoring. . In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.
- Douglas, S.A. (1995). LingWorlds: An intelligent object-oriented environment for second language tutoring. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.
- Driskell, J.E., Willis, R. & Cooper, C. (1992). Effect of overlearning on retention. Journal of Applied Psychology, 77, 615-622.
- Ellis, R. (1986). Understanding second language acquisition. Oxford: Oxford University Press.
- Hamburger, H. (1995). Tutorial tools for language learning by two-medium dialogue. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.
- Kaplan, J.D. & Holland, V.M. (1995). Application of learning principle to the design of a second-language tutor. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.
- Kreyer, S. & Criswell, E. (1995). Instructor as author in an adaptive, multimedia, foreign language tutor. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.
- Lett, J. & O'Mara, F. (1990). Predictors of success in an intensive foreign language context: Correlates of language learning at Defense Language Institute Foreign Language Center. In Perry, T.S. & Stansfield, C.S. (Eds.), Language aptitude reconsidered. Englewood Cliffs, NJ: Prentice Hall.
- Levine, M. (1975). A cognitive theory of learning. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Loftus, G. (1985). Evaluating forgetting curves. Journal of Experimental Psychology (Learning, Memory, and Cognition), 11(2), 397-406.
- Malone, T.W. (1981). Toward a theory of intrinsically motivating instruction. Cognitive Science, 4, 333-369.
- Miller, N. (1963). Some reflections on the law of effect produce a new alternative to drive reduction. In M. R. Jones (Ed.), Nebraska symposium on motivation. Lincoln: University of Nebraska Press.
- Oxford, R.L. (1995) Linking theories of learning with intelligent computer-assisted language learning (ICALL). In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.
- Sams, M.R. (1995). Advanced technologies for language learning: The BRIDGE project within the ARI language tutor program. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Malwah, NJ: Lawrence Erlbaum Associates.

- Schmidt, R.A. & Bjork, R. (1992). New conceptualizations of practice. Psychological Science, 3(4), 207-217.
- Schoelles, M., & Hamburger, H. (1996). Cognitive tools for language pedagogy. Computer Assisted Language Learning, 9(2-3), 213-234.
- Skinner, B.F. (1957). Verbal behavior. New York: Appleton-Century-Crofts.
- Weinberg, A., Garman, J., Martin, J., & Merlo, P. (1995). A principle-based parser for foreign language tutoring in German and Arabic. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), Intelligent language tutors: Theory shaping technology. Mahwah, NJ: Lawrence Erlbaum Associates.